

On uniform convergence in ergodic theorems for a class of skew product transformations

J. BRETTSCHEIDER

Abstract

Consider a class of skew product transformations consisting of an ergodic or a periodic transformation on a probability space (M, \mathcal{B}, μ) in the base and a semigroup of transformations on another probability space (Ω, \mathcal{F}, P) in the fibre. Under suitable mixing conditions for the fibre transformation, we show that the properties ergodicity, weakly mixing, and strongly mixing are passed on from the base transformation to the skew product (with respect to the product measure). We derive ergodic theorems with respect to the skew product on the product space.

The main aim of this paper is to establish uniform convergence with respect to the base variable for the series of ergodic averages of a function F on $M \times \Omega$ along the orbits of such a skew product. Assuming a certain growth condition for the coupling function, a strong mixing condition on the fibre transformation, and continuity and integrability conditions for F , we prove uniform convergence in the base and $L^p(P)$ -convergence in the fibre. Under an equicontinuity assumption on F we further show P -almost sure convergence in the fibre. Our work has an application in information theory: It implies convergence of the averages of functions on random fields restricted to parts of stair climbing patterns defined by a direction.

1 Introduction

The approximation of a line by a planar lattice yields a stair climbing pattern. Consider the averages of a function of a random field along a finite window moving up the stair climbing pattern. Under which conditions does this sequence converge, and what are explicit formulae for the limit? More formally, let $L_{\lambda,t}(z) := (z, [\lambda z + t])$ ($z \in \mathbb{Z}$) be a lattice approximation of the line with slope λ and y -intercept t , and let $m \in \mathbb{N}$ be a fixed window size. Let P be a \mathbb{Z}^2 -indexed random field with values in a set Υ , i.e., a stationary probability measure on $\Omega := \Upsilon^{\mathbb{Z}^2}$, and let \mathcal{F} be the canonical σ -algebra. Consider the averages

$$\frac{1}{n} \sum_{i=0}^{n-1} f(\omega(L_{\lambda,t}(i, \dots, I + m - 1))) \quad (n \in \mathbb{N}) \quad (1)$$

of a function $f \in L^1(\Omega, \mathcal{F}, P)$. What can we say about P -almost sure or $L^1(P)$ -convergence of this sequence? Averages of this type are similar to the ones that occur in the context of directional Shannon-MacMillan theorems for lattice random fields (cf. [6]).

The situation described above can be represented as a special case of a more general set up involving skew product transformations. In the independent component, or *base*, we have a measure-preserving transformation τ on a probability space (M, \mathcal{B}, μ) . In the dependent component, or *fibre*, we have a mixing semigroup of measure-preserving transformations $(\theta_k)_{k \in K}$ on a probability space (Ω, \mathcal{F}, P) , which is linked to the base by a K -valued \mathcal{B} -measurable function κ on M . The class of skew products considered in this paper is given by

$$S(t, \omega) = (\tau(t), \theta_{\kappa(t)}\omega) \quad (t \in M, \omega \in \Omega). \quad (2)$$

This paper deals with the following two questions:

- (A) Is S ergodic or mixing with respect to the product measure?
 - (B) Do the ergodic averages along the skew product converge uniformly with respect to the base?
- Under suitable assumptions, we will answer both questions positively.

Ergodicity, and other mixing properties of various classes of skew products have been studied by a number of authors, but the above situation does not fit into any of the settings covered by existing literature. Kakutani [15] introduced a skew product with a Bernoulli-shift in the base and an ergodic transformation in the fibre was introduced by . He showed that the skew product is ergodic if and only if the transformation in the fibre is ergodic. Other mixing properties were investigated, e.g., by Meilijson [18], den Hollander and Keane [8], and Georgii [12]. Adler and Shields (cf. [1] and [2]) considered a translation on the circle for the fibre. Anzai [3] introduced skew products of two translations on the circle, and derived a criterion for ergodicity. Furstenberg [10] studied unique ergodicity. Zhang [23] investigated this for a translation on a torus in the fibre. A torus translations in the base can also be combined with the translation on \mathbb{R} by the value of a real function of the argument in the base. Skew products of this type are called real extensions of torus translations, and they were explored in Oren [20], Hellekalek and Larcher [13], [14], and Pask [21].

Our answer to question (A) is summarized in Theorem 2.2. We prove that, under suitable mixing conditions for the transformations in the fibre, the properties ergodic, weakly mixing, and strongly mixing, are passed on from the transformation in the base to the skew product. As an explicit example we study the case when P is a random field and $(\theta_k)_{k \in K}$ is a group of shift transformations. In this case, the conditions on the fibre transformation can be insured by assuming tail-triviality for P and a growth condition for ergodic sums of κ along τ (cf. Corollary 2.3).

Our answer to question (B) is given in Theorem 4.9. The proof combines two different approaches. The first approach explores uniform convergence theorems in the spirit of Weyl's classical result for the rotation on the circle. As a little addition to the theorems of Weyl and Oxtoby we show that the ergodic averages of a continuous and uniquely ergodic transformation on a real interval convergence uniformly for the class of Riemann-integrable functions (cf. Corollary 4.4). The second approach are techniques developed for ergodic theorems along subsequences. We extend Blum and Hanson's theorem to the d -parameter case replacing the strict monotonicity condition on the sequence by a growth condition on the coupling function κ , uniformly in t . Combining the two approaches we obtain uniform convergence in the base and L^p -convergence in the fibre for functions that are continuous with respect to the base and fulfill an integrability condition in the fibre (cf. Theorem 4.9). We further derive a variation of this theorem for the class of Riemann-integrable functions on a real interval (cf. Corollary 4.10). We further derive a result (cf. Theorem 4.11) about uniform convergence in the base and P -almost sure convergence in the fibre, provided the iterates of the function fulfill an equicontinuity condition.

We conclude the paper by returning to our initial questions about the asymptotics of (1). Let $\mathbb{T} := [0, 1)$ be the circle equipped with the Borel σ -algebra \mathcal{B} and the Lebesgue measure μ . For $\lambda \in \mathbb{R}$ define a rotation on \mathbb{T} by $\tau_\lambda(t) := t + \lambda \bmod 1$. For $x \in \mathbb{R}$ let $[x]$ be the integer part of x . Define a skew product on the product space $\mathbb{T} \times \Omega$ by

$$S(t, \omega) := (\tau_\lambda(t), \vartheta_{(1, [\lambda+t])}\omega) \quad (t \in \mathbb{T}, \omega \in \Omega). \quad (3)$$

We will see that the iterates of S follow the stair climbing pattern $L_{\lambda, t}$. This allows to rewrite the sequence (1) as an average of f along the orbit of S . The convergence of this sequence, for *all* starting levels t of the stair climbing pattern, is a consequence of the uniform ergodic theorems derived in in Section 3 (cf. Corollary 4.12).

Outline of the paper: In the first section we define mixing properties of semigroups of transformations along sequences and we introduce a the class of skew products considered in this paper. In Theorem 2.2, we give the result on ergodicity and mixing properties of these skew products. Finally, we have a closer look on the case when the fibre transformation is a shift operator for a random field. In Section 3 we discuss ergodic theorems for skew products with ergodic base transformation (cf. Corollary 3.1) and with periodic base transformations (cf. Corollary 3.3). We illustrate the results with two examples related to the sequence (1). In the last section we focus on the main aim of this paper, the uniform convergence with respect to the base. Depending, among other things, on the regularity of the function with respect to the base variable, we obtain different

kinds of convergence in the fibre. Theorem 4.9 states $L^p(P)$ -convergence provided the function is continuous with respect to the base variable. Corollary 4.10 is a version of this for Riemann-integrable functions. Theorem 4.11 states P -almost sure convergence, provided the functions fulfill a certain equicontinuity condition. Corollary 4.12 brings us back to the original motivation for this paper. It states the convergence of the sequences (1).

2 Mixing properties of a class of skew products

Let (Ω, \mathcal{F}, P) be a probability space, and let $(\theta_k)_{k \in \mathbb{N}_0^d}$ be a d -parameter semigroup of measure-preserving transformations on (Ω, \mathcal{F}, P) , i.e., each of the transformations preserves the measure P , and $\theta_0 = \text{Id}$, and $\theta_k \circ \theta_l = \theta_{k+l}$ for all $k, l \in \mathbb{N}_0^d$. $(\theta_k)_{k \in \mathbb{Z}_0^d}$ is a d -parameter group if $(\theta_k)_{k \in \mathbb{N}_0^d}$ is a semigroup and $\theta_{-k} = \theta_k^{-1}$ for all $k \in \mathbb{N}^d$. The following example will be used frequently in our settings. Let σ_1 and σ_2 be two commuting measure-preserving transformations on (Ω, \mathcal{F}, P) . Then

$$\theta_k := \sigma_1^{k^{(1)}} \circ \sigma_2^{k^{(2)}} \quad \text{for } k = (k^{(1)}, k^{(2)}) \in \mathbb{N}_0^2 \quad (4)$$

defines a two-parameter semigroup $(\theta_k)_{k \in \mathbb{N}_0^2}$ of measure-preserving transformations on (Ω, \mathcal{F}, P) . If σ_1 and σ_2 are invertible it extends to a two-parameter group $(\theta_k)_{k \in \mathbb{Z}^2}$. The construction extends to d -parameters in an obvious way.

Let $K = \mathbb{N}_0^d$ or $K = \mathbb{Z}^d$. Let τ be a measure-preserving transformation of a probability space (M, \mathcal{B}, μ) , and assume that κ is a \mathcal{B} -measurable function on M with values in K . Then

$$S(t, \omega) = (\tau(t), \theta_{\kappa(t)}\omega) \quad (t \in M, \omega \in \Omega)$$

defines a skew product on the product space $\overline{\Omega} := M \times \Omega$. In particular, choosing $\kappa \equiv k_0$ for a constant $k_0 \in K$ yields the uncoupled product of τ and θ_{k_0} . Obviously, S is measurable with respect to the product σ -algebra $\mathcal{F} := \mathcal{B} \otimes \mathcal{F}$, and it preserves the product measure $\overline{P} := \mu \otimes P$. It is easy to see that for all $n, m \in \mathbb{N}_0$, for all $t \in M$, and for all $\omega \in \Omega$,

$$S^n(t, \omega) = (\tau^n(t), \theta_{\kappa_n(t)}\omega) \quad \text{and} \quad \kappa_{n+m} = \kappa_n + \kappa_m \circ \tau^n, \quad \text{where } \kappa_n = \sum_{i=0}^{n-1} \kappa \circ \tau^i. \quad (5)$$

Let us now study under which conditions the skew product is ergodic. Furthermore, as suggested by J. Aaronson, we broaden the question to other mixing properties. Answers to these questions will be given in the next lemma. Note that, by a simple projection argument, the ergodicity of τ is necessary for the ergodicity of S . As we know in the case of an uncoupled product, the ergodicity of both transformations does not guarantee the ergodicity of the product. However, it can be shown that the product is ergodic whenever one of the transformations is ergodic and the other one is weakly mixing (cf. [17]). A major ingredient for S for our lemma are assumptions that bring into play the function κ . We make use of two conditions:

- (C1) $(\theta_k)_{k \in K}$ is weakly mixing along the sequence $(\kappa_n(t))_{n \in \mathbb{N}}$ for μ -almost all $t \in M$.
- (C2) $(\theta_k)_{k \in K}$ is strongly mixing and $(\kappa_n(t))_{n \in \mathbb{N}}$ goes to infinity for μ -almost all $t \in M$.

Note that (C2) implies (C1). Condition (C2) can be easily verified for lattice approximations of a line, as discussed in the context of Corollary 4.12. We further need the notion of *weakly mixing along a sequence*. This has been introduced for transformations by N. Friedman (cf. [9]), and we extend it to d -parameter (semi-)groups.

Definition 2.1. Let $(k_n)_{n \in \mathbb{N}}$ be a K -valued sequence. A (semi-)group $(\theta_k)_{k \in K}$ of measure-preserving transformations on (Ω, \mathcal{F}, P) , is called *weakly mixing along $(k_n)_{n \in \mathbb{N}}$ with respect to P* , if

$$\frac{1}{n} \sum_{i=0}^{n-1} |P(A \cap \theta_{k_i}^{-1} B) - P(A)P(B)| \xrightarrow{n \rightarrow \infty} 0 \quad \text{for all } A, B \in \mathcal{F}. \quad (6)$$

Lemma 2.2. (i) Assume condition (C1). If τ is ergodic w.r.t. μ then S is ergodic w.r.t. \overline{P} .

(ii) Assume condition (C1). If τ is weakly mixing w.r.t. μ then S is weakly mixing w.r.t. \overline{P} .

(iii) Assume condition (C2). If τ is strongly mixing w.r.t. μ then S is strongly mixing w.r.t. \overline{P} .

Proof. We give the proof of the first statement here; the remaining proofs are conducted in a similar fashion. Assume condition (C1) and let τ be ergodic with respect to μ . To prove the ergodicity of S we will show that for all bounded $\overline{\mathcal{F}}$ -measurable functions F and G

$$\frac{1}{n} \sum_{i=1}^n \int_{M \times \Omega} F \circ S^i(t, \omega) \cdot G(t, \omega) d\overline{P} \xrightarrow{n \rightarrow \infty} \tilde{F} \cdot \tilde{G},$$

where

$$\tilde{F} = \int_{M \times \Omega} F(t, \omega) d\overline{P} \quad \text{and} \quad \tilde{G} = \int_{M \times \Omega} G(t, \omega) d\overline{P}.$$

It is sufficient to show this for functions which are products of functions on the factors, i.e., $F(t, \omega) = f(t)\Phi(\omega)$ and $G(t, \omega) = g(t)\Psi(\omega)$, where f and g are bounded \mathcal{B} -measurable functions on M and Φ and Ψ are bounded \mathcal{F} -measurable functions on Ω . (The general case follows by approximation.) For these functions we have $\tilde{F} = \tilde{f} \cdot \tilde{\Phi}$ and $\tilde{G} = \tilde{g} \cdot \tilde{\Psi}$, with

$$\tilde{f} = \int_M f d\mu, \quad \tilde{\Phi} = \int_{\Omega} \Phi dP, \quad \tilde{g} = \int_M g d\mu, \quad \text{and} \quad \tilde{\Psi} = \int_{\Omega} \Psi dP,$$

and we obtain

$$\begin{aligned} & \left| \frac{1}{n} \sum_{i=1}^n \int_{M \times \Omega} F \circ S^i(t, \omega) \cdot G(t, \omega) d\overline{P} - \tilde{F} \cdot \tilde{G} \right| \\ & \leq \left| \int_M \frac{1}{n} \sum_{i=1}^n \int_{M \times \Omega} f(\tau^i(t))g(t)\Phi(\theta_{\kappa_i(t)}\omega)\Psi(\omega) - f(\tau^i(t))g(t)\tilde{\Phi}\tilde{\Psi} d\overline{P} \right| d\mu \\ & \quad + \left| \frac{1}{n} \sum_{i=1}^n \left(\int_M f(\tau^i(t))g(t) d\mu - \tilde{f} \cdot \tilde{g} \right) \cdot \tilde{\Phi} \cdot \tilde{\Psi} \right| \\ & \leq \|f\|_{\infty} \|g\|_{\infty} \frac{1}{n} \sum_{i=1}^n \left| \int_{\Omega} \Phi(\theta_{\kappa_i(t)}\omega)\Psi(\omega) dP - \tilde{\Phi} \cdot \tilde{\Psi} \right| + \left| \frac{1}{n} \sum_{i=1}^n \int_M f \circ \tau^i \cdot g d\mu - \tilde{f} \cdot \tilde{g} \right| |\tilde{\Phi}| |\tilde{\Psi}|. \end{aligned}$$

This goes to 0, because the averages in the first expression convergence to 0 for μ -almost all t by condition (C1) and the second expression converges to 0 because of the ergodicity of τ . \square

This section concludes with a closer look at the case of shift transformations on a discrete random field with values in a set Υ . Let $\Omega := \Upsilon^{\mathbb{Z}^d}$. For any $J \subseteq \mathbb{Z}^d$ let \mathcal{F}_J denote the σ -algebra generated by all projections $\omega \mapsto \omega(j)$ with $j \in J$, and let $\mathcal{F} := \mathcal{F}_{\mathbb{Z}^d}$. Denote the coordinates of an elements in \mathbb{Z}^d by upper indices, and let $\|\cdot\|$ be its maximum norm. Consider the *shift transformations* $(\vartheta_v)_{v \in \mathbb{Z}^d}$ on Ω , i.e., $\vartheta_v(\omega)(j) := \omega(j+v)$ ($j \in \mathbb{Z}^d$). Let P be a *random field*, i.e., a measure on (Ω, \mathcal{F}) which is invariant with respect to ϑ_v for all $v \in \mathbb{Z}^d$. The *tail field* is the σ -algebra $\mathcal{T} := \bigcap_{V \subset \mathbb{Z}^d \text{ finite}} \mathcal{F}_{\mathbb{Z}^d \setminus V}$, and P is called *tail-trivial* if it fulfills a 0-1 law on \mathcal{T} . Let $v_1, v_2 \in \mathbb{Z}^d$. As in (4), $\theta_k := \vartheta_{v_1}^{k^{(1)}} \circ \vartheta_{v_2}^{k^{(2)}}$ ($k \in \mathbb{Z}^d$) defines a 2-parameter group of measure-preserving transformations. We have

$$\theta_{\kappa_n(t)} = \vartheta_{v_1}^{\kappa_n^{(1)}(t)} \circ \vartheta_{v_2}^{\kappa_n^{(2)}(t)} = \vartheta_{\kappa_n^{(1)}(t)v_1 + \kappa_n^{(2)}(t)v_2}. \quad (7)$$

In this situation we have the following

Corollary 2.3. Let v_1 and v_2 be linear independent vectors in \mathbb{Z}^d . Assume that P is tail-trivial and that the sequence $(\|\kappa_n(t)\|)_{n \in \mathbb{N}}$ goes to infinity for μ -almost all $t \in M$. Then, when τ is ergodic, weakly mixing or strongly mixing with respect to μ , S is ergodic, weakly mixing or strongly mixing with respect to \overline{P} , respectively.

Proof. We are going to show condition (C2). Define the boxes $V_n = \{v \in \mathbb{Z}^d \mid \|v\| \leq n\}$ ($n \in \mathbb{N}$), and let $B \in \mathcal{F}_J$, for some finite subset J of \mathbb{Z}^d . Then there is an $m \in \mathbb{N}$ such that $J \subseteq V_m$. Setting $m(n) := \kappa_n^{(1)}(t) v_1 + \kappa_n^{(2)}(t) v_2$ we observe that the translated sets $J - m(n)$ are contained in $V_{\tilde{m}(n)}^c$, where $\tilde{m}(n) = (m(n) - 2m) \vee 0$. For any $A \in \mathcal{F}$, we obtain

$$\begin{aligned} |P(A \cap \theta_{\kappa_n(t)}^{-1} B) - P(A) P(B)| &= |P(A \cap \vartheta_{\kappa_n^{(1)}(t) v_1 + \kappa_n^{(2)}(t) v_2}^{-1} B) - P(A) P(B)| \\ &\leq \sup_{C \in \mathcal{F}_{\mathbb{Z}^d \setminus V_{\tilde{m}(n)}}} |P(A \cap C) - P(A) P(C)|. \end{aligned}$$

By the assumptions on v_1, v_2 and κ , $\|\tilde{m}(n)\|$ goes to infinity. By Proposition 7.9 in [11], tail-triviality is equivalent to *short-range correlations*, i.e., $\sup_{C \in \mathcal{F}_{\mathbb{Z}^d \setminus V_n}} |P(A \cap C) - P(A) P(C)|$ converges to 0 as n goes to infinity. \square

3 Ergodic theorems with skew products

Applying Birkhoff's ergodic theorems to the skew product S yields, for any $F \in L^1(\overline{\Omega}, \overline{\mathcal{F}}, \overline{P})$,

$$\frac{1}{n} \sum_{i=0}^{n-1} F(\tau^i(t), \theta_{\kappa_i(t)} \omega) \xrightarrow{n \rightarrow \infty} \overline{E}[F | \mathcal{J}] \quad \overline{P}\text{-almost surely and in } L^1(\overline{P}), \quad (8)$$

where \mathcal{J} is the σ -algebra of all S -invariant sets in $\overline{\mathcal{F}}$. We study this limit more closely for two different cases: when the transformation τ is ergodic and when it is periodic. In the ergodic case, combining (8) and Lemma 2.2 immediately yields the following ergodic theorem for the skew product.

Corollary 3.1. *Assume that τ is ergodic with respect to μ and that the condition (C1) is fulfilled. Then for any function $F \in L^1(\overline{\Omega}, \overline{\mathcal{F}}, \overline{P})$,*

$$\frac{1}{n} \sum_{i=0}^{n-1} F(\tau^i(t), \theta_{\kappa_i(t)} \omega) \xrightarrow{n \rightarrow \infty} \overline{E}[F] \quad \overline{P}\text{-almost surely and in } L^1(\overline{P}).$$

Now consider the case that τ is periodic. We calculate the iterates of the skew product and derive an ergodic theorem with an explicit expression for the limit.

Lemma 3.2. *Assume that τ is periodic with $q \in \mathbb{N}$. Then for all $j \in \mathbb{Z}$ and all $\nu \in \{0, 1, \dots, q-1\}$,*

- (i) $\kappa_{jq+\nu} = j\kappa_q + \kappa_\nu$,
- (ii) $\theta_{\kappa_{jq+\nu}(t)} = \left(\theta_{\kappa_q(t)}\right)^j \circ \theta_{\kappa_\nu(t)} \quad \text{for all } t \in M$,
- (iii) $S^{jq+\nu}(t, \omega) = (\tau^\nu(t), \theta_{\kappa_\nu(t)} \circ (\theta_{\kappa_q(t)})^j \omega) \quad \text{for all } t \in M \text{ and all } \omega \in \Omega$.

Proof. (i) follows from the definition of κ using the periodicity of τ . (ii) is an immediate consequence of (i) and the semigroup property of θ . (iii) follows from (5), the periodicity of τ and because of (ii). \square

Corollary 3.3. *Assume τ is periodic with period $q \in \mathbb{N}$, and $F \in L^1(\overline{\Omega}, \overline{\mathcal{F}}, \overline{P})$. Denote by \mathcal{J}_t the σ -algebra of $\theta_{\kappa_q(t)}$ -invariant sets in \mathcal{F} . Then*

$$\frac{1}{n} \sum_{i=0}^{n-1} F(\tau^i(t), \theta_{\kappa_i(t)} \omega) \xrightarrow{n \rightarrow \infty} \frac{1}{q} \sum_{\nu=0}^{q-1} E[F(\tau^\nu(t), \theta_{\kappa_\nu(t)}(\cdot)) | \mathcal{J}_t],$$

for μ -almost all $t \in M$, and for P -almost all $\omega \in \Omega$ and in $L^1(P)$. If $\theta_{\kappa_q(t)}$ is ergodic with respect to P for μ -almost all $t \in M$, then the limit simplifies to the constant $1/q \sum_{\nu=0}^{q-1} E[F(\tau^\nu(t), \cdot)]$.

Proof. Any $n \in \mathbb{N}$ can be represented as $n = mq + \nu$, with $m \in \mathbb{N}$ and $\nu \in \{0, 1, \dots, q-1\}$, and we may break down the ergodic averages to

$$A_n F := \frac{1}{n} \sum_{i=0}^{n-1} F \circ S^i = \frac{mq}{mq + \nu} \left(\frac{1}{mq} \sum_{i=0}^{mq-1} F \circ S^i + \frac{1}{mq} \sum_{i=mq}^{mq+\nu-1} F \circ S^i \right).$$

Since the first factor converges to 1, and the second addend within the brackets converges to 0, our question reduces to the study of ergodic limits along the subsequence $(mq)_{m \in \mathbb{N}}$. They take the form $A_{mq} F = 1/q \sum_{\nu=0}^{q-1} A_m^{(\nu)} F$, where

$$A_m^{(\nu)} F(t, \omega) := \frac{1}{m} \sum_{j=0}^{m-1} F \circ S^{jq+\nu}(t, \omega) = \frac{1}{m} \sum_{j=0}^{m-1} F(\tau^\nu(t), \theta^{\kappa_\nu(t)} \circ (\theta^{\kappa_q(t)})^j \omega).$$

The last equality can be seen by applying Lemma 3.2. For μ -almost all $t \in M$, the function $f_t^{(\nu)}(\omega) := F(\tau^\nu(t), \theta_{\kappa_\nu(t)} \omega)$ ($\omega \in \Omega$) is integrable, and applying Birkhoff's ergodic theorem yields

$$\lim_{m \rightarrow \infty} A_m^{(\nu)} F(t, \cdot) = E[f_t^{(\nu)} | \mathcal{J}_t] = E[F(\tau^\nu(t), \theta_{\kappa_\nu(t)}(\cdot)) | \mathcal{J}_t] \quad P\text{-almost surely and in } L^1(P)$$

This implies the first statement of the Corollary. In the ergodic case \mathcal{J}_t is trivial, and, using the invariance of P under θ , the last expression reduces to $E[F(\tau^\nu(t), \cdot)]$. \square

We end this section by illustrating the results by two special cases relevant to the skew product tracing a stair climbing pattern introduced in (3).

Example 3.4. *Circle rotations as base transformations.* Let $\mathbb{T} := [0, 1)$ be the circle equipped with the Borel σ -algebra \mathcal{B} and the Lebesgue measure μ . For $\lambda \in \mathbb{R}$ define a rotation on \mathbb{T} by $\tau_\lambda(t) := t + \lambda \bmod 1$. The rotation preserves the measure μ and it is continuous. For rational λ the rotation is periodic, for irrational λ it is uniquely ergodic. The circle can be equipped with a metric $d(s, t) := |s - t|$ ($s, t \in \mathbb{T}$). The metric is rotation invariant.

(i) Let λ be irrational. Choose $(\theta_k)_{k \in K}$ and κ fulfilling condition (C1). By Lemma 2.2, S is ergodic. By Corollary 3.1, for any integrable function F on $(\mathbb{T} \times \Omega, \mathcal{B} \otimes \mathcal{F}, \mu \otimes P)$,

$$\frac{1}{n} \sum_{i=0}^{n-1} F(t + i\lambda \bmod 1, \theta_{\kappa_i(t)} \omega) \xrightarrow{n \rightarrow \infty} \int_0^1 E[F(t, \cdot)] dt,$$

for $\mu \otimes P$ -almost all $(t, \omega) \in \mathbb{T} \times \Omega$ and in $L^1(\mathbb{T} \times \Omega, \mathcal{B} \otimes \mathcal{F}, \mu \otimes P)$.

(ii) Let λ be rational. There is a unique representation $\lambda = p/q$, where $p \in \mathbb{Z}, q \in \mathbb{N}$, p and q have no common divisor. τ_λ is periodic with period q . Furthermore, τ_λ respects the partition $[0, \frac{1}{q}), [\frac{1}{q}, \frac{2}{q}), \dots, [\frac{q-1}{q}, 1)$ of \mathbb{T} , i.e., for every $\nu \in \{1, \dots, q-1\}$ there is a $\tilde{\nu} \in \{1, \dots, q-1\}$ such that $\tau_\lambda([\frac{\nu-1}{q}, \frac{\nu}{q})) = [\frac{\tilde{\nu}-1}{q}, \frac{\tilde{\nu}}{q})$. The limit in Corollary 3.3 is of the form $1/q \sum_{\nu=0}^{q-1} E[F(t + \nu\lambda \bmod 1, \cdot) | \mathcal{J}_t]$. If P is ergodic with respect to $\theta_{\kappa_q(t)}$ then limit simplifies to $1/q \sum_{\nu=0}^{q-1} E[F(t + \nu\lambda \bmod 1, \cdot)]$.

Example 3.5. *Shifts as fibre transformations.* Consider the 2-parameter group define above (7).

(i) If τ is ergodic then, for any function $F \in L^1(\overline{\Omega}, \overline{\mathcal{F}}, \overline{P})$,

$$\frac{1}{n} \sum_{i=0}^{n-1} F(\tau^i(t), \vartheta_{\kappa_i^{(1)}(t) v_1 + \kappa_i^{(2)}(t) v_2}) \xrightarrow{n \rightarrow \infty} \overline{E}[F] \quad \overline{P}\text{-almost surely and in } L^1(\overline{P}).$$

(ii) If τ is periodic with period q then, for any function $F \in L^1(\overline{\Omega}, \overline{\mathcal{F}}, \overline{P})$,

$$\frac{1}{n} \sum_{i=0}^{n-1} F(\tau^i(t), \vartheta_{\kappa_i^{(1)}(t) v_1 + \kappa_i^{(2)}(t) v_2} \omega) \xrightarrow{n \rightarrow \infty} \frac{1}{q} \sum_{\nu=0}^{q-1} E[F(\tau^\nu(t), \vartheta_{\kappa_\nu^{(1)}(t) v_1 + \kappa_\nu^{(2)}(t) v_2}(\cdot)) | \mathcal{J}_t],$$

for μ -almost all $t \in M$, and for P -almost all $\omega \in \Omega$ and in $L^1(P)$. If $\vartheta_{\kappa_q^{(1)}(t)v_1 + \kappa_q^{(2)}(t)v_2}$ is ergodic with respect to P , for μ -almost all $t \in M$, then the limit simplifies to $1/q \sum_{\nu=0}^{q-1} E[F(\tau^\nu(t), \cdot)]$.

4 Uniform convergence

This section addresses the question of *sure* convergence with respect to the first parameter. In addition to the assumptions at the beginning of Section 2 we suppose that M is a compact separable metric space endowed with metric d , and \mathcal{B} is the Borel σ -algebra on M for the topology induced by d . Recall that the convergence of ergodic averages need not be true *everywhere*, even if we are in a compact topological space and both the transformation and the function are continuous. Which conditions guarantee sure convergence in the first parameter? We will be asking a little more than this, namely about *uniform* convergence in t . We are interested in results of the type

$$\frac{1}{n} \sum_{i=0}^{n-1} F(\tau^i(t), \theta_{\kappa_i(t)}\omega) \xrightarrow{n \rightarrow \infty} \overline{E}[F|\mathcal{J}](t, \omega) \quad \text{uniformly in } t \in M \text{ and in } L^1(P). \quad (9)$$

We further investigate when (9) takes place P -almost surely, i.e., for P -almost all $\omega \in \Omega$,

$$\lim_{n \rightarrow \infty} \sup_{t \in M} \left| \frac{1}{n} \sum_{i=0}^{n-1} F(\tau^i(t), \theta_{\kappa_i(t)}\omega) - \overline{E}[F|\mathcal{J}](t, \omega) \right| = 0. \quad (10)$$

Again, we consider two different cases: when τ is ergodic and when it is periodic. In the second case we proceed as in the proof of Corollary 3.3 and obtain the following uniform version.

Corollary 4.1. *Let τ be periodic with $q \in \mathbb{N}$. Assume $F \in L^1(\overline{\Omega}, \overline{\mathcal{F}}, \overline{P})$ with $F(t, \cdot) \in L^1(\Omega, \mathcal{F}, P)$ for all $t \in M$. Then, for P -almost all ω and in $L^1(P)$,*

$$\frac{1}{n} \sum_{i=0}^{n-1} F(\tau^i(t), \theta_{\kappa_i(t)} \cdot) \xrightarrow{n \rightarrow \infty} \frac{1}{q} \sum_{\nu=0}^{q-1} E[F(\tau^\nu(t), \theta_{\nu_q(t)} \cdot) | \mathcal{J}_t] \quad \text{uniformly in } t \in M,$$

where \mathcal{J}_t denotes the σ -algebra of $\theta_{\kappa_q(t)}$ -invariant sets in \mathcal{F} . If P is ergodic with respect to $\theta_{\kappa_q(t)}$, for all $t \in M$, the limit equals $1/q \sum_{\nu=0}^{q-1} E[F(\tau^\nu(t), \cdot)]$.

The ergodic case is more delicate. We begin with careful investigations of ergodic theorems on the single spaces, which will later be combined to derive a result on the product space. In the base we are dealing with a transformation on a compact and metrizable space. We now recall and refine some of the existing results about uniform convergence in this situation. Motivated by applications in information theory, we put particular emphasis on extending uniform convergence results to the class of Riemann-integrable functions. An example for a function that is Riemann-integrable but not continuous occurs in the proof of a directional Shannon-MacMillan theorem for random fields (cf. [6]).

The classical example for an ergodic theorem that gives a statement about uniform convergence is the one by Weyl. In its simplest form, it says that the averages of a continuous function along the orbit of an irrational translation on the circle converge uniformly to the integral of the function. To prove Weyl's theorem, Krengel (cf. Theorem 2.6 in Paragraph 1.2.3 in [17]) uses an Arzela-Ascoli technique which we will make use of at the end of this section. Under the assumption that $\tau : M \rightarrow M$ is continuous and that f is a function on M , such that the functions $F_n := 1/n \sum_{i=1}^{n-1} F \circ \tau^i$ ($n \in \mathbb{N}$) are equicontinuous on M , Krengel's theorem states that the convergence in Birkhoff's ergodic theorem is uniform in t . Together with the following Lemma, it yields Weyl's theorem.

Lemma 4.2. *Let f be a continuous function on M , and $\tau : M \rightarrow M$ Lipschitz-continuous with Lipschitz constant $c \leq 1$. Then the functions F_n ($n \in \mathbb{N}$) defined above are equicontinuous.*

Proof. We have to show that for every $\varepsilon > 0$ there is a $\delta > 0$ such that for all $n \in \mathbb{N}$ and all $s, t \in M$ with $d(s, t) < \delta$, $1/n \left| \sum_{i=0}^{n-1} f(\tau^i(s)) - f(\tau^i(t)) \right| < \varepsilon$. Fix $\varepsilon > 0$. We will show that there is a $\delta > 0$, such that $|f(\tau^i(s)) - f(\tau^i(t))| < \varepsilon$ for all $d(s, t) < \delta$. Since M is compact, f must be uniformly continuous, i.e., there is a $\delta > 0$ such that for all $x, y \in M$ with $d(x, y) < \delta$ we have $|f(x) - f(y)| < \varepsilon$. By assumption, $d(\tau(s), \tau(t)) \leq c d(s, t)$ for all $s, t \in M$, and therefore, $d(\tau^i(s), \tau^i(t)) \leq c^i d(s, t) \leq d(s, t)$ for all $s, t \in M$, and for all $i \in \mathbb{N}$. \square

We shall ask whether we could replace the assumption of continuity of the function f in Weyl's theorem by a weaker condition. It is certainly not true for all measurable functions, which can be seen in a simple example: Fix $t_0 \in \mathbb{T}$. Its orbit under τ is the set $\mathcal{O} := \{\tau^n(t_0) | n \in \mathbb{N}_0\}$. For the function $f := 1_{\mathcal{O}}$, the ergodic averages converge to 1, for all $t \in \mathcal{O}$, but $\int_0^1 1_{\mathcal{O}}(t) dt = 0$.

This question of uniform convergence has a connection with unique ergodicity (cf. e.g., Chapter 4.1.e. of [16] or Theorem 6.19 in [22]). A continuous transformation τ of a compact metrizable space is called uniquely ergodic if it has only one invariant Borel measure. It can be shown that this measure must be ergodic, which implies that the ergodic averages of an integrable function converge almost surely to a constant. The Lebesgue measure is the only probability measure on $(\mathbb{T}, \mathcal{B})$, which is invariant with respect to rotations of the circle, so it must be uniquely ergodic. Oxtoby extended Weyl's theorem to the situation where τ is a uniquely ergodic transformation on a compact metric space. It states uniform convergence of the ergodic averages of a continuous function. Note that, conversely, uniform convergence does not imply the continuity of the function. (Further conditions for this would be needed, such as topological transitivity of τ or constancy of the limit.)

Below Theorem 2.7 in Chapter 1 of [17], Krengel mentions that Weyl's theorem is sometimes spelled out for to the class of Riemann-integrable functions. Actually, it was proved by de Bruijn and Post [7] that the function is Riemann-integrable if and only if the convergence is uniform. This also follows from our next proposition. We ask the following question: Considering uniform convergence of the ergodic averages along a continuous transformation on a compact real interval, can we pass automatically from the class of continuous functions to the class of functions which are integrable in the sense of Riemann?

Proposition 4.3. *Let $a, b \in \mathbb{R}$, $a < b$, μ a measure on $([a, b], \mathcal{B}([a, b]))$ which is absolutely continuous with respect to Lebesgue measure, with a continuous density. Let $\tau : [a, b] \rightarrow [a, b]$ be continuous. Assume that for any continuous function f on $[a, b]$*

$$\frac{1}{n} \sum_{i=0}^{n-1} f \circ \tau^i \xrightarrow{n \rightarrow \infty} \int_a^b f d\mu \quad \text{uniformly.} \quad (11)$$

Then the convergence holds as well for any function which is integrable in the sense of Riemann.

The proof is carried out using a common sandwich argument (cf. e.g., in Chapter 4.1.e. of [16]). Applying the proposition to the situation of Oxtoby's Theorem yields

Corollary 4.4. *Let $a, b \in \mathbb{R}$, $a < b$, and $\tau : [a, b] \rightarrow [a, b]$ continuous and uniquely ergodic with invariant measure μ . Assume that μ is absolutely continuous with respect to Lebesgue measure, with a continuous density. Then, for any function f on M which is integrable in the sense of Riemann,*

$$\frac{1}{n} \sum_{i=0}^{n-1} f \circ \tau^i \xrightarrow{n \rightarrow \infty} \int_0^1 f d\mu \quad \text{uniformly.}$$

Now we focus on studying the convergence in the fibre. Fix $t \in M$, and define a function on Ω by $f(\omega) := F(t, \omega)$. This reduces the ergodic averages of the skew product to $1/n \sum_{i=0}^{n-1} f(\theta_{\kappa_i(t)} \omega)$,

which we can view as a sort of ergodic average along the subsequence $(\kappa_i(t))_{i \in \mathbb{N}}$. Recalling a classical result about L^p -convergence of ergodic averages along subsequences, there is the following characterization.

Theorem 4.5. (Blum & Hanson) *Let T be a transformation on (Ω, \mathcal{F}) . Suppose that T is invertible and that both, T and T^{-1} preserve P . Then P is strongly mixing with respect to T if and only if for all p , $1 \leq p < \infty$, every strictly increasing sequence $(m_i)_{i \in \mathbb{N}}$ of integers, and every function $f \in L^p(\Omega, \mathcal{F}, P)$,*

$$\frac{1}{n} \sum_{i=0}^{n-1} f \circ T^{m_i} \xrightarrow{n \rightarrow \infty} E[f] \quad \text{in } L^p(P).$$

The key to the proof of Blum and Hanson's theorem is the following

Lemma 4.6. *Under the assumptions of Theorem 4.5 and supposing that P is strongly mixing with respect to T we have for all $A \in \mathcal{F}$, for every strictly increasing sequence $(k_i)_{i \in \mathbb{N}}$,*

$$\frac{1}{n} \sum_{i=0}^{n-1} 1_A \circ T^{k_i} \xrightarrow{n \rightarrow \infty} P(A) \quad \text{in } L^2(P).$$

Our next step is to carry over Blum and Hanson's theorem to the case of a d -parameter group of transformations $(\theta_k)_{k \in \mathbb{Z}^d}$, at the place of iterates of one transformation T . What we need is a condition on the \mathbb{Z}^d -valued sequence $(k_i)_{i \in \mathbb{N}}$ which replaces the strict monotonicity imposed on $(m_i)_{i \in \mathbb{N}}$. With an eye toward later applications on the product space we generalize the result further by showing that the L^2 -convergence takes place uniformly over a family of functions, indexed by a set I . Recall that $\|\cdot\|$ denotes the maximum norm in \mathbb{Z}^d .

Lemma 4.7. *Assume that P is strongly mixing with respect to $(\theta_k)_{k \in \mathbb{Z}^d}$, and let $(k_n(t))_{n \in \mathbb{N}}$ ($t \in I$) be a family of sequences with values in \mathbb{Z}^d that fulfill, for all $m \in \mathbb{N}$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sup_{t \in I} |\{1 \leq i, j \leq n \mid \|k_i(t) - k_j(t)\| \leq m\}| = 0. \quad (12)$$

Then for all $A \in \mathcal{F}$,

$$\sup_{t \in I} \left\| \frac{1}{n} \sum_{i=0}^{n-1} 1_A \circ \theta_{k_i(t)} - P(A) \right\|_{L^2(P)}^2 \xrightarrow{n \rightarrow \infty} 0.$$

Proof. For every $A \in \mathcal{F}$ and $t \in I$ we obtain by simple calculations,

$$\begin{aligned} \left\| \frac{1}{n} \sum_{i=0}^{n-1} 1_A \circ \theta_{k_i(t)} - P(A) \right\|_{L^2(P)}^2 &= \int_{\Omega} \frac{1}{n^2} \sum_{i,j=0}^{n-1} (1_A \circ \theta_{k_i(t)} - P(A))(1_A \circ \theta_{k_j(t)} - P(A)) dP \\ &= \frac{1}{n^2} \sum_{i,j=0}^{n-1} \left[\int_{\Omega} (1_A \circ \theta_{k_i(t)} 1_A \circ \theta_{k_j(t)}) dP - P(A) \int_{\Omega} (1_A \circ \theta_{k_i(t)} + 1_A \circ \theta_{k_j(t)}) dP + P(A)^2 \right] \\ &= \frac{1}{n^2} \sum_{i,j=0}^{n-1} \left(P(\theta_{k_i(t)}^{-1} A \cap \theta_{k_j(t)}^{-1} A) - P(A)^2 \right) \leq \frac{1}{n^2} \sum_{i,j=0}^{n-1} \left| P(\theta_{k_i(t)-k_j(t)}^{-1} A \cap A) - P(A)^2 \right|. \end{aligned} \quad (13)$$

Fix $\varepsilon > 0$. Due to the mixing condition there is an $m \in \mathbb{N}$ such that

$$\left| P(\theta_{k_i(t)-k_j(t)}^{-1} A \cap A) - P(A)^2 \right| < \frac{\varepsilon}{2} \quad \text{for all } i, j \in \mathbb{N} \text{ with } \|k_i(t) - k_j(t)\| > m,$$

and

$$\left| P(\theta_k^{-1} A \cap A) - P(A)^2 \right| < \frac{\varepsilon}{2} \quad \text{for all } k \in \mathbb{Z}^2 \text{ with } \|k\| \geq m.$$

By assumption (12) there is a $n_0 \in \mathbb{N}$ such that

$$\frac{1}{n^2} \sup_{t \in I} |\{1 \leq i, j \leq n \mid \|k_i(t) - k_j(t)\| \leq m\}| < \frac{\varepsilon}{2} \quad \text{for all } n \geq n_0.$$

Applying the last two inequalities to (13) yields for all $n \geq n_0$

$$\sup_{t \in I} \left\| \frac{1}{n^2} \sum_{i=0}^{n-1} 1_A \circ \theta_{k_i(t)} - P(A) \right\|_{\mathbb{L}^2(\Omega, \mathcal{F}, P)}^2 < \varepsilon,$$

and the assertion of the lemma follows by letting ε to 0. \square

Theorem 4.8. *Assume that P is strongly mixing with respect to $(\theta_k)_{k \in \mathbb{Z}^d}$. Let $(k_n(t))_{n \in \mathbb{N}}$ ($t \in I$) be a family of sequences with values in \mathbb{Z}^d , for which for all $m \in \mathbb{N}$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sup_{t \in I} |\{1 \leq i, j \leq n \mid \|k_i(t) - k_j(t)\| \leq m\}| = 0.$$

Then for $1 \leq p < \infty$ and for any $f \in L^p(\Omega, \mathcal{F}, P)$,

$$\frac{1}{n} \sum_{i=0}^{n-1} f \circ \theta_{k_i(t)} \xrightarrow{n \rightarrow \infty} E[f] \quad \text{in } L^p(P), \text{ uniformly in } t \in I. \quad (14)$$

Proof. As an immediate consequence of the preceding lemma, (14) is true for $p = 2$ for any simple function g on (Ω, \mathcal{F}) . By a standard argument (cf., e.g., Lemma 4 in [5]), this convergence holds as well in $L^p(P)$, for $1 \leq p < \infty$. Finally, for any function f in $L^p(P)$, decomposition into positive and negative parts, $L^p(P)$ -approximation by simple functions, and monotone convergence yields (14). \square

We will now combine the approaches developed separately for the base and the fibre transformation to derive a result about uniform convergence in the base and L^p -convergence in the fibre. A crucial ingredient is a condition that regulates the effects of the coupling sequence $(\kappa_n(t))_{n \in \mathbb{N}}$.

Theorem 4.9. *Let $\tau : M \rightarrow M$ be continuous and uniquely ergodic, and suppose that P is strongly mixing with respect to the group of transformations $(\theta_k)_{k \in \mathbb{Z}^d}$. Let $\kappa : M \rightarrow \mathbb{Z}^d$ be \mathcal{B} -measurable such that,*

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sup_{t \in M} |\{1 \leq i, j \leq n \mid \|\kappa_i(t) - \kappa_j(t)\| \leq m\}| = 0 \quad (15)$$

for all $m \in \mathbb{N}$. Let be $1 \leq p < \infty$. Then for every \mathcal{F} -measurable function F on $\overline{\Omega}$ such that $\sup_{t \in M} |F(t, \cdot)|$ is in $L^p(\Omega, \mathcal{F}, P)$ and $F(\cdot, \omega)$ is continuous on M for P -almost every ω ,

$$\frac{1}{n} \sum_{i=0}^{n-1} F(\tau^i(t), \theta_{\kappa_i(t)} \cdot) \xrightarrow{n \rightarrow \infty} \overline{E}[F] \quad \text{in } L^p(P) \text{ and uniformly in } t \in M. \quad (16)$$

Proof. We first prove the theorem in the case when F is the indicator of a set of the form $U \times A$, where U is the intersection of finitely many metric balls in M or their complements, and $A \in \mathcal{F}$. By (5), the expression

$$\left\| \frac{1}{n} \sum_{i=0}^{n-1} F(\tau^i(t), \theta_{\kappa_i(t)} \cdot) - \overline{E}[F] \right\|_{L^2(P)} \quad (17)$$

then transforms to

$$\begin{aligned} & \frac{1}{n^2} \sum_{i,j=0}^{n-1} \int_{\Omega} 1_U(\tau^i(t)) 1_U(\tau^j(t)) 1_A(\theta_{\kappa_i(t)} \omega) 1_A(\theta_{\kappa_j(t)} \omega) P(d\omega) \\ & - \frac{1}{n^2} \sum_{i,j=0}^{n-1} \mu(U) P(A) \left[\int_{\Omega} 1_U(\tau^i(t)) 1_A(\theta_{\kappa_i(t)} \omega) P(d\omega) + \int_{\Omega} 1_U(\tau^j(t)) 1_A(\theta_{\kappa_j(t)} \omega) P(d\omega) \right] \\ & + \mu(U)^2 P(A)^2. \end{aligned} \quad (18)$$

By $P(\theta_{\kappa_i(t)}^{-1}A \cap \theta_{\kappa_j(t)}^{-1}A) = P(\theta_{\kappa_i(t)-\kappa_j(t)}^{-1}A \cap A)$, the first addend equals

$$\frac{1}{n^2} \sum_{i,j=0}^{n-1} 1_U(\tau^i(t))1_U(\tau^j(t))P(\theta_{\kappa_i(t)-\kappa_j(t)}^{-1}A \cap A).$$

It may be replaced by

$$\frac{1}{n^2} \sum_{i,j=0}^{n-1} 1_U(\tau^i(t))1_U(\tau^j(t))P(A)^2 \quad (19)$$

without affecting the asymptotic behavior of (18) as can be seen as follows. We may bound

$$\begin{aligned} & \left| \frac{1}{n^2} \sum_{i,j=0}^{n-1} 1_U(\tau^i(t))1_U(\tau^j(t)) \left(P(\theta_{\kappa_i(t)-\kappa_j(t)}^{-1}A \cap A) - P(A)^2 \right) \right| \\ & \leq \frac{1}{n^2} \sum_{i,j=0}^{n-1} \left| P(\theta_{\kappa_i(t)-\kappa_j(t)}^{-1}A \cap A) - P(A)^2 \right|. \end{aligned}$$

Now, we argue as in the second part of the proof of Lemma 4.7, replacing the sequence $(k_i)_{i \in \mathbb{N}}$ by $(\kappa_i(t))_{i \in \mathbb{N}}$, and using assumption (15) instead of (12). This proves that the difference created by the change (19) converges to 0 uniformly with respect to t .

Since the term in the rectangular brackets in the second addend in (18) equals $1_U(\tau^i(t))P(A) + 1_U(\tau^j(t))P(A)$, the whole expression (18) simplifies to

$$\frac{1}{n^2} \sum_{i,j=0}^{n-1} \left(1_U(\tau^i(t))1_U(\tau^j(t)) - \mu(U)(1_U(\tau^i(t)) + 1_U(\tau^j(t))) + \mu(U)^2 \right) P(A)^2,$$

which can be further reduced to $\left(1/n \sum_{i=0}^{n-1} 1_U(\tau^i(t)) - \mu(U) \right)^2 P(A)^2$. Since τ is uniquely ergodic and $\mu(\partial U) = 0$, Corollary 4.1.14 in [16] tells us that the first factor converges to 0 uniformly in t , which concludes the first part of the proof. To pass from L^2 -convergence to general L^p , use again a standard argument (for instance, Lemma 4 in [5]).

Now we let F be a general function, satisfying the conditions of the theorem. We need to find for every positive ϵ a sequence of metric balls U_i in M and $A_i \in \mathcal{F}$, with real numbers a_i such that, for all $t \in M$, $\|F(t, \cdot) - I(t, \cdot)\|_p < \epsilon$, where $I = \sum_{i=1}^n a_i 1_{U_i \times A_i}$. It will then follow that

$$\left\| \frac{1}{n} \sum_{i=0}^{n-1} F(\tau_i(t), \theta_{\kappa_i(t)} \cdot) - \frac{1}{n} \sum_{i=0}^{n-1} I(\tau_i(t), \theta_{\kappa_i(t)} \cdot) \right\|_p \leq \frac{1}{n} \sum_{i=0}^{n-1} \|F(\tau_i(t), \cdot) - I(\tau_i(t), \cdot)\|_p < \epsilon.$$

For $\omega \in \Omega$ and $c > 0$, let $\delta(c, \omega)$ be the modulus of continuity for the function $F(\cdot, \omega)$. Define the sets

$$M_k = \sup_{t \in M} \{\omega \mid |F(t, \omega)| \leq k\} \quad \text{and} \quad D_k(\epsilon) = \{\omega \mid \delta(1/k, \omega) \leq \epsilon\} \quad (k \in \mathbb{N}).$$

Then the sequence of functions on Ω , $F_k(\omega) := \sup_{t \in M} |F|^p(t, \omega) 1_{D_k(\epsilon/4)^c \cup M_k^c}(\omega)$ ($k \in \mathbb{N}$), is bounded by $\sup_{t \in M} |F|^p(t, \omega)$, which is integrable, and converges to 0 for every ω . By the bounded convergence theorem, the integral of F_k converges to 0 as k goes to infinity. Choose a k such that $\int F_k P(d\omega) \leq (\epsilon/2)^p$. Since M is compact, we may find a finite sequence $t_1, \dots, t_r \in M$ such that the balls of radius $1/k$ around these centers cover M . We also define a sequence of real numbers $-k-1 = s_0 < \dots < s_{r'} = k$ such that the difference between any two successive elements is less than $\epsilon/8$. Now we define a collection of sets $U_{i,j}$ and $A_{i,j}$ indexed by $r \times r'$. We start with $U_{i,j}$ as the ball of radius $1/k$ around t_i , and then remove the intersections, so that the $U_{i,j}$ is the same for all j , and running through $1 \leq i \leq r$ yields a disjoint cover of M . The sets $A_{i,j}$ are defined by

$$A_{i,j} = \{\omega \mid s_{j-1} < F(t_i, \omega) \leq s_j\} \cap D_k(\epsilon/8) \cap M_k.$$

Let $a_{i,j} = s_j$. We throw in one additional product set, $U_0 = M$ and $A_0 = D_k(\varepsilon/8)^c \cup M_k^c$ with $a_0 = 0$, and define the simple function $I(t, \omega)$ as indicated above. Then for any $t \in M$,

$$\left\| F(t, \cdot) - I(t, \cdot) \right\|_p = \left(\int |F(t, \omega) - I(t, \omega)|^p 1_{D_k(\varepsilon/8)} 1_{M_k} P(d\omega) \right)^{1/p} + \left(\int F_k(\omega) P(d\omega) \right)^{1/p}.$$

We already assumed (in defining k) that the second term is smaller than $\varepsilon/2$. For every t , there is a unique pair i, j such that $t \in U_{i,j}$ and $\omega \in A_{i,j}$. By construction, $I(t, \omega) = s_j$, so the integrand in the first term is bounded by $2^p |F(t_i, \omega) - F(t, \omega)|^p + 2^p |F(t_i, \omega) - s_j|^p$. This in turn is bounded by $2^p(\varepsilon/4)^p < (\varepsilon/2)^p$, since ω is not in A_0 and $d(t_i, t) < 1/k$, completing the proof. \square

Proposition 4.3 yields the following version for Riemann-integrable functions.

Corollary 4.10. *Let $a, b \in \mathbb{R}$, $a < b$, and $\tau : [a, b] \rightarrow [a, b]$ be continuous and uniquely ergodic with invariant measure μ , and assume that μ is absolutely continuous with respect to Lebesgue measure, with continuous density. For P and κ assume the same as in Theorem 4.9. Let $F \in L^p([a, b] \times \Omega, \mathcal{B} \otimes \mathcal{F}, \overline{P})$ be Riemann-integrable with respect to the first variable. Then we have*

$$\lim_{n \rightarrow \infty} \sup_{t \in M} \left\| \frac{1}{n} \sum_{i=0}^{n-1} F(\tau^i(t), \theta_{\kappa_i(t)} \cdot) - \overline{E}[F] \right\|_{L^p(P)} = 0.$$

Our next goal is to derive a statement about P -almost sure convergence rather than $L^1(P)$ -convergence in the fibre. Further conditions on F are needed. P -almost sure convergence of ergodic theorems involving weights or subsequences is a very subtle question (cf., e.g. [4]). Choosing a function F which is constant in ω and considering Lemma 4.2 suggests that we need an equicontinuity assumption in t . Note also the additional assumptions that non-empty open sets on M have positive mass under μ .

Theorem 4.11. *Let μ be a τ -invariant measure on (M, \mathcal{B}) , such that any non-empty open subset U of M has $\mu(U) > 0$. Let $\tau : M \rightarrow M$ be continuous and F a function on $M \times \Omega$, for which $F(t, \cdot) \in L^1(P)$ for all $t \in M$, and the sequence of functions*

$$\left(\frac{1}{n} \sum_{i=0}^{n-1} F(\tau^i(\cdot), \theta_{\kappa_i(\cdot)} \omega) \right)_{n \in \mathbb{N}}$$

is equicontinuous on M , for all $\omega \in \Omega$. Then, for P -almost all $\omega \in \Omega$,

$$\sup_{t \in M} \left| \frac{1}{n} \sum_{i=0}^{n-1} F(\tau^i(t), \theta_{\kappa_i(t)} \omega) - \overline{E}[F|\mathcal{J}](t, \omega) \right| \xrightarrow{n \rightarrow \infty} 0. \quad (20)$$

Proof. We may assume without loss of generality that $E[F|\mathcal{J}] = 0$. The general case can be reduced to this by subtracting $E[F|\mathcal{J}]$ on both sides and making use of the invariance of $E[F|\mathcal{J}]$ under S . The first step is to construct a countable dense set $M_1 \subset M$ and a set $N_1 \subset \Omega$ with $P(N_1) = 0$ such that

$$\frac{1}{n} \sum_{i=0}^{n-1} F \circ S^i(t, \omega) \longrightarrow 0 \quad \text{for all } t \in M_1 \text{ and all } \omega \in \Omega \setminus N_1. \quad (21)$$

Since M is compact, the conditions on F assure that $F \in L^1(\overline{\Omega}, \overline{\mathcal{F}}, \overline{P})$, and therefore by (8) there is a set $M_1 \subset M$ with $\mu(M_1) = 1$ such that for any $t \in M_1$ there is a set $N(t) \subset \Omega$ with $P(N(t)) = 0$ and (21) holds for all $\omega \in \Omega \setminus N(t)$. \widetilde{M}_1 is dense in M because its complement has measure zero with respect to μ and therefore, by assumption, contains no non-empty open subsets. Since M is separable we can find a countable dense subset $C \subset M$, and because M_1 is dense in

M , we can approximate any $x \in C$ by a sequence $(a_j(x))_{j \in \mathbb{N}}$ with $a_j(x) \in \widetilde{M}_1$ for all $j \in \mathbb{N}$. $M_1 := \bigcup_{x \in C} \bigcup_{j \in \mathbb{N}} a_j(x)$ defines a countable dense subset of M , and $N_1 := \bigcup_{t \in M_1} N(t)$ defines a subset of Ω , which fulfills (21). This completes the first step.

For the next step, choose $s \in M$ and fix $\varepsilon > 0$. By equicontinuity, there is a set $N_0 \subset \Omega$ with $P(N_0) = 0$ and a $\delta > 0$ such that for all $r, t \in M$ with $d(r, t) < \delta$

$$\left| \frac{1}{n} \sum_{i=0}^{n-1} F \circ S^i(r, \omega) - F \circ S^i(t, \omega) \right| < \frac{\varepsilon}{2} \quad \text{for all } n \in \mathbb{N} \text{ and all } \omega \in \Omega \setminus N_0. \quad (22)$$

Define $N := N_0 \cup N_1$ and fix $\omega \in \Omega \setminus N$. Since M_1 is dense in M we can find a $t \in M_1$ with $d(s, t) < \delta$, and by (21) there is an $n_1 \in \mathbb{N}$ such that

$$\left| \frac{1}{n} \sum_{i=0}^{n-1} F \circ S^i(t, \omega) \right| < \frac{\varepsilon}{2} \quad \text{for all } n \geq n_1.$$

Combining the last two inequalities leads the desired

$$\left| \frac{1}{n} \sum_{i=0}^{n-1} F \circ S^i(s, \omega) \right| < \varepsilon \quad \text{for all } n \geq n_1 \text{ and all } \omega \in \Omega \setminus N.$$

For uniform convergence w.r.t. the first variable, we use a standard compactness argument. M can be covered by a finite number m of δ -neighborhoods in M , which centers are denoted by s_1, \dots, s_m . Applying the reasoning of the last step to each of the s_1, \dots, s_m we find $n_0 \in \mathbb{N}$ such that

$$\left| \frac{1}{n} \sum_{i=0}^{n-1} F \circ S^i(s_k, \omega) \right| < \frac{\varepsilon}{2} \quad \text{for all } n \geq n_0, k \in \{1, \dots, K\}, \text{ and } \omega \in \Omega \setminus N.$$

For an arbitrary $s \in M$ there exists $k \in \{1, \dots, K\}$ such that $d(s, s_k) < \delta$, and by (22) we obtain

$$\frac{1}{n} \left| \sum_{i=0}^{n-1} F \circ S^i(s, \omega) - F \circ S^i(s_k, \omega) \right| < \frac{\varepsilon}{2} \quad \text{for all } n \in \mathbb{N} \text{ and all } \omega \in \Omega \setminus N.$$

Finally, the convergence (20) follows by the last two inequalities. \square

We conclude the paper with an application of our results to the question that originally motivated this work. Consider the ergodic averages of a function of a random field restricted to the staircase pattern defined by the approximation of a line by a lattice as in (1). We actually the function to be multivariate. In this context, this means that it may depend on a finite number of such steps. What can be said about the asymptotic behavior of the ergodic averages of such a function?

Let P be a two-dimensional random field, that is, a probability measure on $\Omega = \Upsilon^{\mathbb{Z}^2}$ invariant w.r.t. the group of shift transformations $(\vartheta_v)_{v \in \mathbb{Z}^2}$ (see just above Corollary 2.3 for details). For a set $U \subseteq \mathbb{Z}^2$ let $\pi_U(\omega) := \omega(U)$ and $P_U := P \circ \pi_U^{-1}$ defines a probability distribution on $\Upsilon^{|U|}$. Use $[x]$ denote the integer part of a real number x , respectively. Recall that $L_{\lambda, t}(z) = (z, [\lambda z + t])$ ($z \in \mathbb{Z}$) is an approximation of the line with slope λ and y -intercept t by elements of the lattice \mathbb{Z}^2 . For $z_1, z_2 \in \mathbb{Z}$, $L_{\lambda, t}(z_1, \dots, z_2)$ are the z_1 th to the z_2 th step. We will use the short form $P_{\lambda, t, m} := P_{L_{\lambda, t}(0, \dots, m-1)}$.

Corollary 4.12. *Let $m \in \mathbb{N}$ and $\lambda \in [0, 1]$. Let f be a function on Υ^m which is integrable w.r.t. $P_{\lambda, t, m}$ for all $t \in [0, 1]$.*

(i) *Let λ be rational and represented as $\lambda = p/q$ for $p \in \mathbb{Z}$ and $q \in \mathbb{N}$ with no common divisor. If P is ergodic w.r.t. the shift transformations then*

$$\frac{1}{n} \sum_{i=0}^{n-1} f(\omega(L_{\lambda, t}(i, \dots, i+m-1))) \xrightarrow{n \rightarrow \infty} \frac{1}{q} \sum_{\nu=0}^{q-1} \int_{\Upsilon^m} f(y) P_{\lambda, \tau^\nu(t), m}(dy)$$

in $L^1(P)$, for P -almost all $\omega \in \Omega$, and uniformly in $t \in [0, 1]$.

(ii) Let λ be irrational. Assume further that $f(\omega(L_{\lambda,t}(0, \dots, m-1)))$ is Riemann-integrable with respect to $t \in [0, 1]$, for P -almost all $\omega \in \Omega$. If P is strongly mixing w.r.t. the shift transformations then

$$\frac{1}{n} \sum_{i=0}^{n-1} f(\omega(L_{\lambda,t}(i, \dots, i+m-1))) \xrightarrow{n \rightarrow \infty} \int_0^1 \int_{\Upsilon^m} f(y) P_{\lambda,t,m}(dy) dt$$

in $L^1(P)$ and uniformly in $t \in [0, 1]$.

Proof. Let τ_λ be the rotation on the circle defined in Example 3.4, and S_λ the skew product defined in (2). Use $\kappa(t) := (1, [t + \lambda])$ and κ_n as defined in (5). It is easy to show that $\kappa_n(t) = L_{\lambda,t}(n)$ for all $n \in \mathbb{N}$: For $n = 1$, it follows immediately from plugging in the definition, and it remains to induce the statement from n to $n+1$. It is obvious for the first coordinate. The second coordinate of κ can be written as $\kappa_{n+1}^{(2)}(t) = \kappa_n^{(2)}(t) + \kappa \circ \tau_\lambda^n(t) = [n\lambda + t] + [\tau_\lambda^n(t) + \lambda]$. The claim now follows from $[\tau_\lambda^n(t) + \lambda] = [t + n\lambda - [t + n\lambda] + \lambda] = -[t + n\lambda] + [t + (n+1)\lambda]$.

This implies that the iterates of the skew product are of the form $S_\lambda^n(t, \omega) = (\tau_\lambda^n(t), \theta_{L_{\lambda,t}(n)}\omega)$ ($n \in \mathbb{N}$) capturing the lattice approximation of the line. Using the second equality in (5), it also follows that $L_{\lambda,t}(n+u) = L_{\lambda,t}(n) + L_{\lambda,\tau^n(t)}(u)$ for all $n, u \in \mathbb{N}_0$. We thus get $L_{\lambda,t}(i, \dots, i+m-1) = L_{\lambda,t}(i) + L_{\lambda,\tau^i(t)}(0, \dots, m-1)$. Introducing the function $F_\lambda(t, \omega) := f(\omega, L_{\lambda,t}(0, \dots, m-1))$ we obtain

$$\begin{aligned} & \frac{1}{n} \sum_{i=0}^{n-1} f(\omega(L_{\lambda,t}(i, \dots, i+m-1))) \\ &= \frac{1}{n} \sum_{i=0}^{n-1} f \circ \vartheta_{L_{\lambda,t}(i)}(\omega(L_{\lambda,\tau^i(t)}(0, \dots, m-1))) = \frac{1}{n} \sum_{i=0}^{n-1} F_\lambda \circ S_\lambda(t, \omega). \end{aligned} \quad (23)$$

Consider case (i). With the representation $\lambda = p/q$ as above, τ_λ is periodic with q . By Corollary 4.1 and (23), the averages on the left side in (i) converge uniformly in t , for P -almost all ω and in $L^1(P)$. Since P is ergodic with respect to $\theta_{\kappa_q(t)}$ for all $t \in M$, the limit equals $1/q \sum_{\nu=0}^{q-1} \int_\Omega F_\lambda(\tau_\lambda^\nu(t), \cdot) P(d\omega)$ which simplifies to the expression on the right hand side of (i).

Consider case (ii). For an irrational λ , τ_λ is uniquely ergodic. Since $\|\kappa_n(t)\|$ (with $\|\cdot\|$ for the maximum norm) is bounded from below by n , the sequence tends to infinity as n goes to infinity. Corollary 2.3 with $v_1 = (1, 0)$ and $v_2 = (0, 1)$ implies the ergodicity of S_λ . It remains to verify condition (15). The latter easily follows from $\|\kappa_i(t) - \kappa_j(t)\| \geq |i - j|$, and since $1/n^2 \{1 \leq i, j \leq n \mid |i - j| \leq m\}$ converges to 0 for all $m \in \mathbb{N}$. Now, Corollary 4.10 applied to (23) implies that the averages in (ii) converge uniformly in t , and in $L^1(P)$. The limit equals $\int_0^1 \int_\Omega F_\lambda(t, \cdot) P(d\omega) dt$ which simplify to the expression on the right hand side of (ii). \square

Note that P -almost everywhere convergence in (ii) can be derived as well, but requires additional conditions of the form stated in Theorem 4.11. For $m = 1$ the limit actually simplifies to the integral of f with respect to the marginal distribution of P in the origin. In particular, it is independent of t , and it is the same in (i) and (ii). The simplest interesting case is $m = 2$. Let $P_{\text{flat}} := P \circ \pi_{\{(0,0),(1,0)\}}^{-1}$ denote the marginal distribution of P on the subset $\{(0,0), (1,0)\}$. This corresponds to the case where $L_{\lambda,t}$ does not have a step in $z = 1$. Let $P_{\text{step}} := P \circ \pi_{\{(0,0),(1,1)\}}^{-1}$ denote the marginal distribution of P on the subset $\{(0,0), (1,1)\}$. This corresponds to the case where $L_{\lambda,t}$ has a step in $z = 1$. Then, the limits in both (i) and (ii) of the above corollary are of the form

$$\lambda \int_{\Upsilon^2} f(y) P_{\text{step}}(dy) + (1 - \lambda) \int_{\Upsilon^2} f(y) P_{\text{flat}}(dy). \quad (24)$$

Acknowledgements

This work was motivated by a special case of Corollary 4.12 which originally occurred during my thesis work. I am grateful to my supervisor Hans Föllmer for the rich education he gave me in both ergodic theory and probability theory. It is a pleasure for me to thank Jon Aaronson for guiding me toward a broader perspective for Lemma 2.2. David Steinsaltz assisted with an approximation argument I used in the proof of Theorem 4.9, and I thank him for this hint. Benjamin Weiss and Yuval Peres helped me situate my results within the field of ergodic theory; I am grateful to them for the interest they showed in my work. My thanks are also due to Hans Crauel, Manfred Denker, Didier Piau, Michael Scheutzow and Masha Saprykina for helpful discussions on different aspects of ergodic theory. I would like to express my gratitude to Bill MacKillop for supporting me in manifold ways during the time I've been working at Queen's University.

References

- [1] R. Adler and P. Shields. Skew products of Bernoulli shifts with rotations. *Israel J. Math.*, 12:215–222, 1972.
- [2] R. Adler and P. Shields. Skew products of Bernoulli shifts with rotations II. *Israel J. Math.*, 19:228–236, 1974.
- [3] H. Anzai. Ergodic skew product transformations on the torus. *Osaka Math. J.*, 3(1), March 1951.
- [4] A. Bellow and V. Losert. The weighted pointwise ergodic theorem and the individual ergodic theorem along subsequences. *Trans. Am. Math. Soc.*, 288:307–345, 1985.
- [5] J.R. Blum and D.L. Hanson. On the mean ergodic theorem for subsequences. *Bull. Am. Math. Soc.*, 66:308–311, 1960.
- [6] J. Brettschneider. Shannon-MacMillan theorems for random fields along curves and lower bounds for surface-order large deviations. *Tentatively accepted by Prob. Th. Rel. Fields*, 2006.
- [7] N. G. de Bruijn and K. A. Post. A remark on uniformly distributed sequences and Riemann integrability. *Indag. Math.*, 30:149–50, 1968.
- [8] F. den Hollander and M. Keane. Ergodic properties of color records. *Physica*, 138A:183–193, 1986.
- [9] N. Friedman. Mixing on sequences. *Can. J. Math.*, 35:339–352, 1983.
- [10] H. Furstenberg. Strict ergodicity and transformation of the torus. *Amer. J. Math.*, 83:573–601, 1961.
- [11] H.-O. Georgii. *Gibbs measures and phase transitions*. W. de Gruyter, Berlin, 1988.
- [12] H.-O. Georgii. Mixing properties of induced random transformations. *Ergod. Th. and Dynam. Systems*, 17:839–847, 1997.
- [13] P. Hellekalek and G. Larcher. On the ergodicity of a class of skew products. *Israel J. Math.*, 54(3):301–306, 1986.
- [14] P. Hellekalek and G. Larcher. On Weyl sums and skew products over irrational rotations. *Theoret. Comput. Sci.*, 65(2):189–196, 1989.
- [15] S. Kakutani. Random ergodic theorems and Markov processes with a stable distribution. In J. Neyman, editor, *Proceedings of the Second Berkeley Symposium on Statistics and Probability*, pages 247–261, Berkeley, 1951. University of California Press.

- [16] A. Katok and B. Hasselblatt. *Introduction to the modern theory of dynamical systems*. Cambridge Univ. Press, Cambridge, 1995.
- [17] U. Krengel. *Ergodic theorems*. W. de Gruyter, Berlin, 1985.
- [18] I. Meilijson. Mixing properties of a class of skew-products. *Israel J. Math.*, 19:266–270, 1974.
- [19] J. Milnor. On the entropy geometry of cellular automata. *Complex Syst.*, 2(3):357–385, 1988.
- [20] I. Oren. Ergodicity of cylinder flows arising from irregularities of distribution. *Israel J. Math.*, 44(2):127–138, 1983.
- [21] D. A. Pask. Skew products over the irrational rotation. *Israel J. Math.*, 69(1):65–74, 1990.
- [22] P. Walters. *An introduction to ergodic theory*. Springer-Verlag, New York, Heidelberg, Berlin, 1982.
- [23] Q. Zhang. On skew products of irrational rotations with tori. In *Convergence in ergodic theory and probability*, volume 5, pages 435–445, 1996.

Julia Brettschneider

Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK

Department of Community Health & Epidemiology and Cancer Research Institute Division of Cancer Care & Epidemiology, Queen's University, Ontario, K7L 3N6, Canada

`julia.brettschneider@warwick.ac.uk`